

Approaches to Evaluating Reproductive Hazards and Risks

by Carole A. Kimmel

Development of approaches to risk assessment for reproductive toxicity has aided in the critical evaluation of the scientific basis for interpretation of data and the description of assumptions underlying the process. In addition, it has helped to standardize, to the extent possible, the use of qualitative and quantitative data in the hazard identification and dose-response processes and to identify research needed to fill gaps in the available database. The standard study protocols for evaluating reproductive and developmental hazards include developmental toxicity studies and both short-term and longer-term reproductive studies. These study protocols have been in use for several decades, but development of risk assessment approaches has prompted the recommendation of additional end point measures to these protocols. These include evaluation of specific neurologic and behavioral function of offspring following prenatal and postnatal exposure, evaluation of sperm production and quality, reproductive organ weights, and more in-depth testicular histopathology in males, as well as measures of age at vaginal opening, vaginal cytology, oocyte toxicity, time to mating, gestation length, and reproductive organ weights in females. Current approaches to risk assessment in reproductive toxicity involve the determination of a no-observed-adverse-effect level (NOAEL) and the application of uncertainty factors (UFs) to account for differences between the experimental animal species and humans, variability in sensitivity within the human population, and other factors as necessary to derive the reference dose (RfD), or a specified RfD for developmental toxicity to account for the short period of exposure required. Recognized limitations to the NOAEL/UF approach for calculating the RfD have prompted the investigation of alternative approaches, such as the benchmark dose approach, which uses more of the data and takes into consideration the slope of the dose-response curve and the variability in the data. Research to improve risk assessment for reproductive toxicity is needed in a number of areas, particularly improved quantitative risk assessment approaches, both in terms of statistical and biologically based modeling approaches.

Introduction

The importance of delineating approaches for analyzing and interpreting data for reproductive toxicity has been realized within the past few years. Although standard studies on the reproductive and developmental effects of chemicals have been conducted and used for regulatory purposes since the early 1960s (and to a limited extent, even earlier), only within the last decade have efforts been made to critically evaluate the scientific basis for interpretation of data and to describe the standard assumptions that are made in the process of risk assessment. This evaluation has also had an impact on the evolution of improved testing approaches and will continue to influence the design of testing protocols as advances are made in

addressing uncertainties when estimating the risk of reproductive toxicity from exposures to humans.

The U.S. Environmental Protection Agency (EPA) has been instrumental in developing risk assessment guidelines for male and female reproductive toxicity, which were proposed in 1988 (1,2), and for developmental toxicity, first published in 1986 (3) with the final publication in 1991 (4). These guidelines are based on the paradigm for risk assessment described originally by the National Research Council (5), which included four components: hazard identification, dose-response assessment, exposure assessment, and risk characterization. Some modifications have been made in the process, which was based primarily on cancer risk assessment, to account for the assumption of a threshold generally made for noncancer health effects, including reproductive and developmental toxicity. Currently, the female and male reproductive guidelines are being combined into one guideline, and publication is expected in 1993.

The risk assessment guidelines define the terms "reproductive toxicity" and "developmental toxicity" (Table 1) and discuss the end points that are considered adverse effects for both human and experimental animal data. Reproduc-

Reproductive and Developmental Toxicology Branch (RD-689), U.S. Environmental Protection Agency, 401 M Street SW, Washington, DC 20460.

This manuscript was presented at the Conference on the Impact of the Environment on Reproductive Health that was held 30 September-4 October 1991 in Copenhagen, Denmark.

Table 1. Risk assessment guidelines: definition of terms.**Reproductive toxicity**

The occurrence of adverse effects on the reproductive system that may result from exposure to environmental agents. Toxicity may be expressed as alterations to the reproductive organs and/or the related endocrine system. The manifestation of such toxicity may include, but not be limited to, alterations in sexual behavior, onset of puberty, fertility, gestation, parturition, lactation, pregnancy outcomes, premature reproductive senescence, or modifications in other functions that are dependent on the integrity of the reproductive system.

Developmental toxicity

The occurrence of adverse effects on the developing organism that may result from exposure before conception (either parent), during prenatal development, or postnatally to the time of sexual maturation. Adverse developmental effects may be detected at any point in the life span of the organism. The major manifestations of developmental toxicity include death of the developing organism, structural abnormality, altered growth, and functional deficiency.

tive toxicity includes developmental toxicity and also refers to effects on the reproductive organs and/or the related endocrine system of males and females. Developmental toxicity includes effects on the developing organism resulting from exposure not only during the prenatal period but also exposure of either parent before mating and postnatal exposure from birth to the time of sexual maturation. Thus, although separate risk assessment guidelines were written for male and female reproductive toxicity and for developmental toxicity because of the complexity of the systems involved, the two are integrally related, such that reproductive effects in adults may result in developmental effects, and effects on the developing reproductive system prenatally or early postnatally may result in reproductive impairment in adulthood. Data on reproduction and development are often found in the same studies (both for human and experimental animal studies) and are usually evaluated in concert.

The primary advantages of developing risk assessment guidance include *a*) explicitly stating the assumptions made in the risk assessment process; *b*) standardizing to the extent possible the use of qualitative and quantitative data in the hazard identification and dose-response processes; and *c*) identifying research needed for reducing uncertainties and to fill gaps in the available database. Each of these points will be discussed in more detail in the following sections.

Assumptions in Reproductive Toxicity Risk Assessment

The EPA risk assessment guidelines discuss the basic assumptions that are generally made in the extrapolation of data from animal studies to humans. Because of many unknowns in the extrapolation process, assumptions must be made on the relevance of effects in animal studies to potential human risk. These assumptions are generally applied in the absence of data but do not preclude further investigation to support or refute the assumptions made. The assumptions are listed in Table 2 and provide the inferential basis for the approaches taken to reproductive risk assessment by the EPA.

Table 2. Basic assumptions for reproductive toxicity risk assessment.

An agent that produces an adverse reproductive effect in experimental animal studies will potentially pose a hazard to humans after sufficient exposure.

Reproductive effects are generally the same across species except for pregnancy outcomes, which may vary depending on species-specific differences in timing of exposure, critical periods, metabolism, developmental patterns, placentation, or mechanism of action.

All of the four manifestations of developmental toxicity (death, structural abnormalities, growth alterations, and functional deficits) are of concern.

The most appropriate species, if known, will be used to estimate human risk; otherwise, the most sensitive species will be used.

A threshold is generally assumed for the dose-response curve for reproductive effects.

First, it is assumed that an agent that produces an adverse reproductive effect in experimental animal studies will potentially pose a hazard to humans following sufficient exposure. This assumption is based on the comparisons of data for known human reproductive toxicants (6-11), which indicate that, in general, experimental animal data are predictive of reproductive effects in humans.

Because the basic male and female reproductive processes are generally similar across species, adverse reproductive effects are assumed generally to be the same across species. In the case of pregnancy outcomes, however, it is assumed that the types of adverse developmental effects seen in animal studies are not necessarily the same as those that may be produced in humans. Every species may not react in the same way to a given agent during development, possibly because of species-specific differences in critical periods, timing of exposure, metabolism, developmental patterns, placentation, or mechanisms of action. Thus, it is difficult to determine which will be the most appropriate species for predicting the specific types of effects seen in humans.

It is assumed that all of the four manifestations of developmental toxicity (death, structural abnormalities, growth alterations, and functional deficits) are of concern. The tendency to consider only malformations or malformations and death as end points of concern ignores the body of data accumulated on the effects of agents on growth alterations and functional deficits in humans and the fact that there is usually at least one experimental species that mimics the types of effects seen in humans.

When sufficient data are available (e.g., pharmacokinetics), it is assumed that the most appropriate species will be used to estimate human reproductive risk. In the absence of such data, the most sensitive species is used, based on observations that humans are often as sensitive or more so than the most sensitive animal species tested.

In general, a threshold is assumed for the dose-response curve for reproductive toxicants. This is based on the known capacity of cells, tissues, and organs of the reproductive system and of the developing organism to compensate for or to repair a certain amount of damage at the cellular, tissue, or organ level. Furthermore, multiple insults at the molecular or cellular level may be required to produce an adverse effect.

Approaches for Evaluating Reproductive Toxicity

Standard Reproductive Toxicity Testing Protocols

Although adequate human data are always preferable for estimating risks of reproductive effects from environmental exposures, data from reproductive toxicity studies in laboratory animal species form the primary database used for risk assessment in this area. The database on reproductive toxicity also may be enhanced by data from other toxicity studies, as well as from pharmacokinetic and mechanistic studies.

Historically, different testing approaches have been used for different types of agents. For pharmaceutical agents, for example, the three-segment design is required for reproductive toxicity testing (12). The segment I study is designed to evaluate fertility and general reproductive function and to assess potential effects on development of offspring. The segment II study is the standard teratology or developmental toxicity study and is designed to provide information on effects from exposure of maternal animals during pregnancy. The segment III study provides an assessment of peri- and postnatal toxicity in exposed dams and their offspring. Recent efforts at international harmonization of pharmaceutical testing guidelines and development of an integrated protocol have resulted in a much more flexible study design depending on the intended use of the drug and other data available (13).

For foods and food additives, a segment II study is required as well as a comprehensive multigeneration study (14) designed to provide information (directly or indirectly) concerning the effects of a test substance on gonadal function, estrous cycles, mating behavior, conception, parturition, pregnancy outcome, lactation, and postnatal growth and viability for up to three generations. For environmental agents, the segment II study and a two-generation reproduction study (15-17) are required.

Recommended Additions to Standard Protocols for Reproductive Toxicity

Data from the multigeneration study provide information on the "couple," since both sexes are treated. Although the studies are not designed specifically to allow determination of the affected sex, evaluation of mating pairs or animals unable to mate successfully may indicate the gender affected. Multigeneration protocols are relatively insensitive in detecting effects on fertility. For example, normal males of most test species produce numbers of sperm that greatly exceed the minimum requirements for fertility as evaluated in current protocols (18-20). Reductions of up to 90% in number of normal sperm may occur without a statistically significant effect on fertility. In humans, however, sperm counts are closer to the threshold for the number of normal sperm needed to ensure full

reproductive competence, and a decrease in number of normal sperm is more likely to result in altered fertility. Therefore, several additions and changes to the basic protocol have been proposed to improve sensitivity and to allow better interpretation and more specific information on the gender(s) affected and on the site of action (e.g., gonad or pituitary). For example, in the male, sperm production and sperm quality, reproductive organ weights, and more in-depth testicular histopathology could be added.

In the female, the difference in sensitivity between rodents and humans may not be as great as for males, and an effect on fertility may reflect changes in the estrous cycle, endocrine function, or oocyte toxicity. A long mating period (up to 3 weeks) allows the possibility of mating over several estrous cycles in the female, and sensitivity for detecting an effect on fertility could be improved by limiting the mating period. The relationship between fertility in females and other measures of reproductive function has not been tested adequately, however, and measures of estrous cycle normality and oocyte toxicity should be included. In addition, potential effects on sexual development and reproductive senescence should be evaluated. Thus, adequate evaluation of female reproductive toxicity should include several measures in addition to those usually obtained, such as age at vaginal opening, vaginal cytology, oocyte toxicity (destruction of the primary oocyte population leading to cessation of ovarian function), time to mating, gestation length, and reproductive organ weights based on stage of the estrous cycle at necropsy.

Another protocol that provides a more sensitive evaluation of subfertility is the "fertility assessment by continuous breeding" protocol conducted in mice or rats (21-23). In this study design, mating pairs are cohabitated for 98 days with continuous exposure. Litters are examined and removed shortly after birth. The number of litters produced, litter size, weight, and any external abnormalities are recorded. The last litter produced is raised to adulthood, exposure is continued, and if an effect has been seen, cross-mating with control offspring is conducted to determine which sex is affected. In addition, numerous reproductive end points are evaluated. With this approach, each pair may deliver up to five litters in the time period designated, and the average number of litters per pair provides an index of fertility or subfertility. This study design takes less time overall than the multigeneration study, but provides additional data including a more sensitive measure of subfertility.

Testing Approaches for Evaluating Functional Developmental Toxicity

The currently available protocols for testing chemicals in laboratory animals provide limited information on the hazards of chemical exposure in neonatal and young animals. In many cases, they could be improved upon to provide more complete information, such as incorporating

tests of functional effects of specific organ systems. The only organ system for which testing guidelines are available in such testing protocols is the central nervous system (CNS).

Although a few countries currently have testing guidelines that call for behavioral testing of offspring, EPA is the only regulatory agency with specific testing guidelines that address a number of issues of protocol design, aspects of CNS function to be included, and criteria for selection of testing procedures. The developmental neurotoxicity protocol (24) for testing pesticides and industrial chemicals was designed to evaluate potential functional and morphological hazards to the developing nervous system that may arise in offspring from exposure of the maternal animal during pregnancy and lactation. It also provides general information on postnatal growth and survival.

Because of its design, the developmental neurotoxicity testing protocol may be conducted as a separate study, concurrently with, or as a follow-up to, a developmental toxicity (segment II) study, or be folded into a multi-generation study in the second generation. It is required on a case-by-case basis depending on what other toxicity information is available on each chemical or class of chemical. Although the developmental neurotoxicity protocol was designed to assess specific effects on the developing nervous system, it could easily be used as a model for evaluating functional and morphological hazards on other organ systems. For example, if an agent is suspected of producing developmental renal toxicity (25,26), the basic framework of this same study design may be used, with possible modification of the period and duration of exposure and substitution of parameters used to assess renal structure and function instead of neurobehavioral effects (27).

Quantitative Approaches to Reproductive Risk Assessment

Current approaches to risk assessment in reproductive toxicity involve the determination of a no-observed-adverse-effect level (NOAEL) from standard studies with a minimal data set (usually three dose levels and a control). Uncertainty factors (UFs) are then applied to account for differences between the experimental animal species and humans, variability in sensitivity within the human population, and other factors as necessary to derive the reference dose (RfD). A specific RfD for developmental toxicity (RfD_{DT}) is determined to account for the short period of exposure required. The RfD is expressed in terms of the exposure duration, route, and timing of exposure. In the case of inhalation exposure, a reference concentration (RfC) is determined. The RfD is assumed to represent a dose at which no excess risk for reproductive effects above background are likely to occur in the human population.

There are several limitations to the NOAEL/UF approach for calculating the RfD: *a*) use of the NOAEL focuses only on the dose that is the NOAEL; in fact, the NOAEL must be one of the experimental doses; *b*) use of the NOAEL ignores the shape of the dose-response curve; *c*) this approach results in smaller studies having higher

NOAELs because data variance is not taken into account; *d*) the NOAEL approach does not result in an estimate of risk at a given dose, especially above the RfD.

The benchmark dose approach was originally proposed in 1984 by Crump (28) as a simple but important improvement in the estimation of the RfD. As shown in Figure 1, the benchmark dose is the lower confidence limit on an effective dose (LED) corresponding to an increase in the incidence of an effect at a particular risk level, e.g., the LED₁₀ is the lower confidence limit on a dose that is effective in producing a 10% increase in response. Uncertainty factors may then be applied to the LED₁₀ to calculate the RfD. Since the NOAEL theoretically can fall anywhere between zero and an incidence just below that detectable as an increase above control levels (usually in the range of 7–10% for quantal data), the benchmark dose would provide a common starting point for applying uncertainty factors and would result in RfDs with more comparable levels of protection than when NOAELs are used. Which benchmark dose to use is still under consideration. Crump (28) and Kimmel and Gaylor (29) discussed the use of the LED₁₀ because it usually falls within the experimental range. If enough data are available at the lower end of the dose-response range, it is also possible to calculate an LED₀₅ or an LED₀₁ [as discussed by Gaylor (30) and Chen and Kodell (31)], values that would be closer to a true NOAEL and that would require application of fewer uncertainty factors than the LED₁₀.

Various mathematical models have been proposed for use in the benchmark dose approach. Theoretically, the choice of the model should not be critical as long as it fits the data well because estimation is within the observed

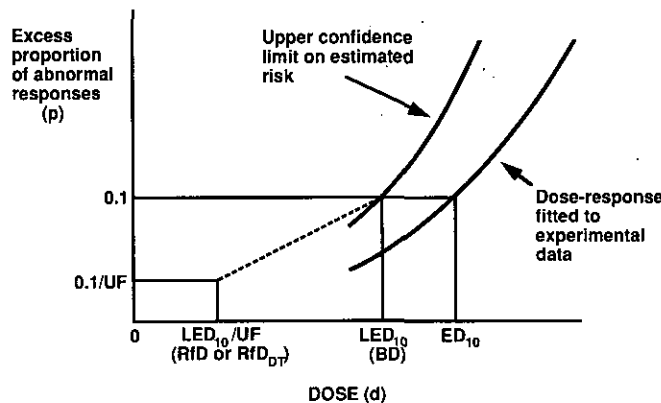


FIGURE 1. Graphic illustration of the benchmark dose approach. The benchmark dose (BD) approach illustrated here [based on Crump (28) and Kimmel and Gaylor (29)] is derived by modeling the data in the observed range, selecting an incidence level within or near the observed range (e.g., the effective dose to produce a 10% increased incidence of response, the ED₁₀), and determining the upper confidence limit on the model. The upper confidence value corresponding to, for example, a 10% excess in response is used to derive the BD, which is the lower confidence limit on dose for that level of excess response, in this case, the LED₁₀. The reference dose (RfD) or RfD for developmental toxicity (RfD_{DT}) estimated by applying uncertainty factors (UF) to the BD would be greater than or equal to the BD/UF.

dose range for most quantal end points. Thus, the assumption of a threshold would not be of concern in the choice of the model because risk would not be extrapolated to low levels of exposure. If, however, there are biological reasons to incorporate particular factors in the model (e.g., intralitter correlations), these should be included to account, as much as possible, for variability in the data.

EPA is currently conducting studies on the application of the benchmark dose to actual data sets contributed by several industrial and government laboratories. Information gained from these efforts will be used to write guidelines for using the benchmark dose approach in the risk assessment process.

The qualitative and quantitative information on hazard and dose-response, along with the NOAEL and RfD, are compared to the human exposure estimates in the final characterization of risk. Risk characterization is the culmination of the risk assessment process, providing an evaluation of the overall quality of the assessment and describing risk in terms of the nature and extent of harm. Table 3 lists the essential components of the risk characterization. A summary of the toxicity information, together with its strengths and weaknesses and the assumptions and uncertainties, is described, and the NOAELs for the various end points of concern (e.g., adult male and female reproductive effects, developmental effects, maternal toxicity), the RfD and RfD_{DT}, the estimates of human exposure, the margin of exposure (estimated human exposure/NOAEL), the overall weight of evidence, and the basis for the risk characterization are given. Several risk characterizations may be appropriate, e.g., based on maximal exposure, average exposure, highly exposed groups, or susceptible subpopulations. This information is then considered along with economic, technological, social, and political factors in deciding how to manage the attendant risks of exposure in the population.

Research Needs in Reproductive Risk Assessment

Research to improve risk assessment is needed in a number of areas, as identified in the risk assessment guidelines and outlined in Table 4. Several of these issues have been explored in workshops supported and/or organized by EPA in which scientists deliberated to reach consensus, where possible, and to identify further

research needs. For example, a workshop held in 1987 (32) focused on the relationship between maternal and developmental toxicity and formed the basis for the current position taken by EPA on this issue, i.e., developmental toxicity in the presence of maternal toxicity is not assumed to be secondary to maternal toxicity (4). However, further research efforts to elucidate the influence of maternal toxicity on the developing offspring and vice versa are needed. Similar issues concerning male and female reproductive toxicity and its relationship to other forms of toxicity also need to be explored.

Another workshop focused on the use of one- versus two-generation studies for reproductive toxicity evaluation (33). Conclusions at this workshop supported the continued use of two-generation reproduction studies to thoroughly evaluate the potential reproductive effects of an agent. More complete characterization and definition of end points for female reproductive toxicity are needed, as is information on the interrelationships among end points of male reproductive toxicity. Such information could improve the sensitivity and predictability of currently used reproductive toxicity protocols.

In 1990, a workshop was held on the qualitative and quantitative comparability of human and animal developmental neurotoxicants (11), at which experts reviewed the human and animal data on known human developmental neurotoxicants. The consensus of opinion at this workshop was that the general types of effects produced by an agent were similar in humans and in animal models and that internal effective dose levels (when data were available) were more similar across species than external exposure levels. A great deal more work is needed on the effects of agents on the CNS and other functional systems, including the critical periods of exposure and improved testing protocols.

Within the last few years, a major focus within the EPA has been on the development of quantitative approaches to

Table 3. Components of a risk characterization.

| |
|--|
| Characterization of the health-related data |
| Range of effective doses |
| No-observed-adverse-effect level (NOAEL) for reproductive and/or developmental effects. |
| Reference dose (RfD) and/or reference dose for developmental toxicity (RfD _{DT}) |
| Assumptions and uncertainties |
| Estimated human exposure |
| Margin of exposure (MOE) |
| Weight of evidence |
| Basis for characterization, e.g., maximal exposure, average exposure, highly exposed groups, sensitive populations |

Table 4. Research needs in reproductive toxicity risk assessment.

| |
|---|
| Male and female reproductive toxicity |
| More completely characterize and define female reproductive end points |
| Develop quantitative extrapolation between end points and function (e.g., sperm count and fertility) |
| Better understand the relationship between reproductive effects and other forms of toxicity |
| Developmental toxicity |
| Further evaluate the relationship between maternal and developmental toxicity |
| Elucidate postnatal functional alterations and critical periods of exposure |
| Evaluate the validity of short-term <i>in vivo</i> and <i>in vitro</i> tests |
| Both |
| Explore the interrelationships among end points |
| Delineate the mechanisms of toxicity and pathogenesis |
| Develop comparative pharmacokinetic data |
| Further examine the threshold concept for dose-response relationships |
| Develop improved mathematical models for dose-response modeling |
| Examine the effects of agents given by various routes of exposure to develop methodology for route-to-route extrapolation |
| Conduct epidemiology studies with more quantitative measures of exposure |

developmental toxicity risk assessment, including the use of the benchmark dose approach and development of biologically based dose-response models (34-36). For the short term, we are studying the application of the benchmark dose methodology to actual data sets, and other projects are focusing on the development of biologically based dose-response models.

One aspect of this work focuses on the interrelationship of multiple outcomes in developmental toxicity studies. In a recent paper (37), we described the process of development, as covered in segment II studies, as a continuum of events leading to resorption or death, or to a viable fetus at term, and for that fetus, its malformation status and weight. Although these manifestations are routinely assessed in segment-II-type studies, the results are typically handled as independent experimental outcomes. Assessment of multiple outcomes is complicated by the presence of competing risks (e.g., implants that die during organogenesis cannot go on to express malformations at term). In addition, it is clear that there are correlations between certain outcomes and that these can be incorporated into models that better characterize the nature of the dose-response relationship. In this study, the joint effect of exposure on fetal weight and malformation status was evaluated because these are two events that can be quantified in an individual fetus. From this evaluation, it was clear that malformed fetuses always tended to be lower in weight than normal fetuses, even those within the same dose group.

Further work is underway on the development of a model incorporating malformations and fetal weight and is the first effort to combine a continuous variable, fetal weight, with a binary variable, the presence or absence of a malformation. This can be accomplished by assuming that there is a latent, continuous variable ultimately involved in the induction of a malformation and that a malformation occurs when the value of the latent variable exceeds some tolerance value for an individual fetus. Intralitter correlations are also important in this relationship, such that fetuses from light litters have a greater chance of being malformed and fetuses that are light with respect to their littermates are even more likely to be malformed. The ultimate aim of this work is to extend the model to include the conditional probability of being live (not resorbed or dead) on fetal weight and malformation status. The advantages of this approach are that if all three outcomes are unrelated, a combined analysis would help ensure an overall effect was not missed, or, if they were correlated, that the resulting analysis would have greater statistical power. Thus, an approach such as this could find immediate application in dose-response modeling for risk assessment.

Summary

It is important to keep in mind the complexity of the sequence of events in the reproductive systems as both testing and risk assessment approaches are reevaluated and advances are made. Despite the problems of extrapolating from animal data to humans and of finding the most

appropriate animal models, it is extremely important to evaluate a broader range of reproductive effects than is now the case in standard protocols. Several additions to current standard protocols have been discussed that could greatly improve the information available for interpretation. In addition, because of the fragmented way in which toxicity testing is done overall, the age-related aspects of various organ or system toxicities have often been overlooked unless specific concerns are raised by observations in nonstandard laboratory animal studies or in exposed human populations. A more logical approach to testing might be to use the two-generation reproduction study approach, with exposure beginning before mating and continuing through pregnancy and lactation, and with observation of various toxicities (e.g., reproductive toxicity, cancer, neurotoxicity, immunotoxicity, etc.) in the resulting offspring. Thus, the two-generation study design could be used or modified to evaluate the age-related aspects of a number of organs or systems. An added bonus is that combining studies results in the use of fewer animals and lower costs for testing than conducting individual studies and may also provide valuable insight into mechanisms of toxicity.

The views expressed in this paper are those of the author and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

REFERENCES

1. U.S. Environmental Protection Agency. Proposed guidelines for assessing female reproductive risk. Fed. Reg. 53: 24834-24847 (1988).
2. U.S. Environmental Protection Agency. Proposed guidelines for assessing male reproductive risk. Fed. Reg. 53: 24850-24869 (1988).
3. U.S. Environmental Protection Agency. Guidelines for the health assessment of suspect developmental toxicants. Fed. Reg. 51: 34028-34040 (1986).
4. U.S. Environmental Protection Agency. Guidelines for developmental toxicity risk assessment; Fed. Reg. 56: 63798-63826 (1991).
5. National Research Council. Risk assessment in the Federal government: managing the process. Committee on the Institutional Means for the Assessment of Risks to Public Health. Commission on Life Sciences, National Research Council. National Academy Press, Washington, DC, 1983, pp. 17-83.
6. Nisbet, I. C. T., and Karch, N. J. Chemical Hazards to Human Reproduction. Noyes Data Corp., Park Ridge, IL, 1983.
7. Kimmel, C. A., Holson, J. F., Hogue, C. J., and Carlo, G. L. Reliability of Experimental Studies for Predicting Hazards to Human Development. NCTR Technical Report No. 6015, National Center for Toxicological Research, Jefferson, AR, 1984.
8. Hemminki, K., and Vineis, P. Extrapolation of the evidence on teratogenicity of chemicals between humans and experimental animals: chemicals other than drugs. *Teratog. Carcinog. Mutagen.* 5: 251-318 (1985).
9. Meistrich, M. L. Critical components of testicular function and sensitivity to disruption. *Biol. Reprod.* 34: 17-28 (1986).
10. Working, P. K. Male reproductive toxicity: Comparison of the human to animal models. *Environ. Health Perspect.* 77: 37-44 (1988).
11. Kimmel, C. A., Rees, D. C., and Francis, E. Z., Eds. Qualitative and quantitative comparability of human and animal developmental neurotoxicity. *Neurotoxicol. Teratol.* 12: 175-292 (1990).
12. FDA. Guidelines for Reproduction and Teratology of Drugs. Bureau of Drugs, Food and Drug Administration, Rockville, MD, 1966.
13. Bass, R. Thoughts on the development of a transnational reproductive toxicity test guideline and presentation of the current proposal. Presented at the Teratology Society Annual Meeting, June 25, Boca Raton, FL, 1991.

14. Food and Drug Administration. Advisory Committee on Protocols for Safety Evaluations. Panel on reproduction report on reproduction studies in the safety evaluation of food additives and pesticide residues. *Toxicol. Appl. Pharmacol.* 16: 264-296 (1970).
15. U.S. Environmental Protection Agency. Pesticides Assessment Guidelines, Subdivision F. Hazard Evaluation: Human and Domestic Animals. EPA-540/9-82-025, National Technical Information Service, Springfield, VA, 1982.
16. U.S. Environmental Protection Agency. Toxic Substances Control Act test guidelines; final rules. *Fed. Reg.* 50: 39426-39428; 39433-39434 (1985).
17. U.S. Environmental Protection Agency. FIFRA Accelerated Reregistration Phase 3 Technical Guidance. EPA 540/09-90-078, National Technical Information Service, Springfield, VA, 1990.
18. Amann, R. P. A critical review of methods for evaluation of spermatogenesis from seminal characteristics. *J. Androl.* 2: 37-58 (1981).
19. Meistrich, M. L. Quantitative correlation between testicular stem cell survival, sperm production, and fertility in the mouse after treatment with different cytotoxic agents. *J. Androl.* 3: 58-68 (1982).
20. Robaire, B., Smith, S., Hales, B. F. Suppression of spermatogenesis by testosterone in adult male rats: effect on fertility, pregnancy outcome and progeny. *Biol. Reprod.* 31: 221-230 (1984).
21. Lamb, J. C., IV. Reproductive toxicity testing: evaluation and developing new test systems. *J. Am. Coll. Toxicol.* 4: 163-171 (1985).
22. Morrissey, R. E., Lamb, J. C., IV, Morris, R. W., Chapin, R. E., Gulati, D. K., and Heindel, J. J. Results and evaluations of 48 continuous breeding reproduction studies conducted in mice. *Fundam. Appl. Toxicol.* 13: 747-777 (1989).
23. Gulati, D. K., Hope, E., Teague, J., and Chapin, R. E. Reproductive toxicity risk assessment by continuous breeding in Sprague-Dawley rats: a comparison of two study designs. *Fundam. Appl. Toxicol.* 17: 270-279 (1991).
24. U.S. Environmental Protection Agency. Pesticide Assessment Guidelines, Subdivision F. Hazard Evaluation: Human and Domestic Animals, Addendum 10, Neurotoxicity, Series 81, 82, 83. No. PB 91-154617, National Technical Information Service, Springfield, VA, 1991.
25. Kavlock, R. J., Rehnberg, B. F., and Rogers, E. H. Renal functional teratogenesis resulting from adriamycin exposure. *Teratology* 33: 213-220 (1986).
26. Kavlock, R. J., Rehnberg, B. F., and Rogers, E. H. Critical prenatal periods for chlorambucil-induced functional alterations of the rat kidney. *Toxicology* 43: 51-64 (1987).
27. Kimmel, C. A., Kavlock, R. J., and Francis, E. Z. Animal models for assessing developmental toxicity. In: *Similarities and Differences Between Children and Adults: Implications for Risk Assessment* (P. S. Guzelian, C. J. Henry, and S. S. Olin, Eds.), ILSI Press, Washington, DC, 1992, pp. 43-65.
28. Crump, K. S. A new method for determining allowable daily intakes. *Fundam. Appl. Toxicol.* 4: 854-871 (1984).
29. Kimmel, C. A., and Gaylor, D. W. Issues in qualitative and quantitative risk analysis for developmental toxicology. *Risk Anal.* 8: 15-20 (1988).
30. Gaylor, D. W. Quantitative risk analysis for quantal reproductive and developmental effects. *Environ. Health Perspect.* 79: 243-246 (1989).
31. Chen, J. J., and Kodell, R. L. Quantitative risk assessment for teratological effects. *J. Am. Stat. Assoc.* 84: 966-971 (1989).
32. Kimmel, G. L., Kimmel, C. A., and Francis, E. Z., Eds. Evaluation of maternal and developmental toxicity. *Teratog. Carcinog. Mutagen.* 7: 203-338 (1987).
33. Francis, E. Z., and Kimmel, G. L. Proceedings of the workshop on one- vs two-generation reproductive effects studies. *J. Am. Coll. Toxicol.* 7: 911-925 (1988).
34. Kimmel, C. A., Wellington, D. G., Farland, W., Ross, P., Manson, J. M., Chernoff, N., Young, J. F., Selevan, S. G., Kaplan, N., Chen, C., Chitlik, L. D., Siegel-Scott, C. L., Valaoras, G., and Wells, S. Overview of a workshop on quantitative models for developmental toxicity risk assessment. *Environ. Health Perspect.* 79: 209-215 (1989).
35. Kimmel, C. A. Dose response modelling approaches for risk assessment. Presented at the Toxicology Forum, July 17, Aspen, CO, 1991.
36. Kavlock, R. J. Biologically based dose-response models for risk assessment. Presented at the Toxicology Forum, July 17, Aspen, CO, 1991.
37. Ryan, L. M., Catalano, P. J., Kimmel, C. A., and Kimmel, G. L. Relationship between fetal weight and malformation in developmental toxicity studies. *Teratology* 44: 215-223 (1991).